

B. ESCOFIER

B. LE ROUX

Influence d'un élément sur les facteurs en analyse des correspondances

Les cahiers de l'analyse des données, tome 1, n° 3 (1976),
p. 297-318

http://www.numdam.org/item?id=CAD_1976__1_3_297_0

© Les cahiers de l'analyse des données, Dunod, 1976, tous droits réservés.

L'accès aux archives de la revue « Les cahiers de l'analyse des données » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

INFLUENCE D'UN ÉLÉMENT SUR LES FACTEURS EN ANALYSE DES CORRESPONDANCES

[INFLUENCE]

par B. Escofier ⁽¹⁾ et B. Le Roux ⁽²⁾

1. Introduction.

Dans l'interprétation des résultats d'une analyse de correspondances, on se demande souvent quelle est l'influence de tel ou tel élément particulier sur un facteur. Les éléments pour lesquels on se pose de telles questions sont ceux qui ont une contribution importante pour ce facteur. On est même quelquefois amené à refaire une analyse en supprimant un tel élément " α " pour avoir une base d'interprétation plus solide. Que peut-il se passer pour le facteur de même rang de la nouvelle analyse ? Divers cas sont possibles.

1) Le facteur est le même, à peu près, que dans l'analyse comportant l'élément α . Il y a stabilité, l'élément incriminé n'a pas une grande importance dans la détermination du facteur. Ce qui ne signifie pas qu'il doit être négligé dans l'interprétation du facteur, au contraire : si un élément de ce type est très corrélé avec un facteur, le facteur traduit bien une tendance générale des données; et cet élément permet de caractériser ou de résumer cette tendance. Bien entendu, pour faire cette interprétation, il faut que le poids (fréquence) de l'élément soit assez élevé pour que sa position soit assez sûre.

2) Le facteur est différent, mais le plan engendré par ce facteur et le facteur associé à la valeur propre immédiatement supérieure, (ou inférieure) est stable. Ceci se produit quand deux valeurs propres sont très proches. Le plan qui leur est associé est très stable, mais il peut y avoir rotation voire permutation de ces deux facteurs dans ce plan. Généralement le plan est à considérer globalement, on dira alors qu'il y a stabilité comme dans le premier cas.

3) Le facteur étudié se retrouve dans les facteurs interprétables de l'analyse sans α , mais son rang est décalé; ceci signifie précisément que, si ce facteur a rang s dans l'analyse complète, il est très proche d'un facteur interprétable de rang supérieur t de l'analyse sans α . Si les facteurs de rang $s + 1, \dots, t$ de l'analyse complète sont à peu près ceux de rang $s, \dots, t - 1$ de l'analyse sans α , la présence de α ne modifie pas essentiellement la forme générale du nuage, elle ne fait qu'augmenter l'inertie du s -ième facteur. Mais on ne parlera pas d'influence de α sur les facteurs de rang s, \dots, t , leur montée est seulement la conséquence de la descente du s -ième facteur.

4) Le facteur disparaît sans qu'il y ait modification des autres facteurs considérés comme significatifs. Il était alors déterminé par

(1) Maître-Assistant I.N.S.A. - Rennes.

(2) Maître-Assistant U.E.R. MLFI - Université René Descartes Paris.

l'élément a . Il n'avait donc pas d'intérêt dans l'analyse globale des données, puisqu'il ne représente pas une tendance générale mais seulement la tendance associée à cet élément.

5) Les facteurs de rang $s, s + 1, \text{etc.}$ sont très différents dans les deux analyses. L'élément a détenait une grande importance dans la détermination de ces facteurs. Avec quelques précautions, on peut souvent dire que a perturbait l'analyse.

Bien entendu, ces cinq cas forment une présentation très schématique des différences possibles entre les résultats des deux analyses. Il n'y a jamais stabilité totale, décalage ou disparition d'un facteur sans aucune perturbation des autres facteurs retenus comme significatifs, mais on est souvent dans une situation plus ou moins proche de l'un de ces cas. Les méthodes que nous présentons ici permettent souvent, en utilisant seulement les résultats de l'analyse du tableau complet, de reconnaître l'une des quatre premières situations. On mesure la stabilité des facteurs par un critère angulaire simple défini au § 2. Nous exposons, au § 3, les méthodes qui permettent, sans faire de nouvelles analyses, d'étudier l'influence d'un élément et qui sont des supports appréciables pour l'interprétation des facteurs; nous donnons sans démonstrations (publiées dans [4]), des formules de majoration dont, pour faciliter l'usage, nous avons tracé des abaques. Le § 4 est un guide pour l'utilisateur lui permettant d'étudier rapidement à l'aide des abaques l'influence d'un élément sur la stabilité des facteurs. Un exemple d'application est donné au § 5.

2. Angle entre facteurs ou plans factoriels issus des deux analyses.

Pour étudier l'influence d'un élément a , on compare donc les facteurs issus de l'analyse du tableau complet $I \times J$, à ceux issus du tableau $I \times (J - \{a\})$ où on a supprimé la colonne a

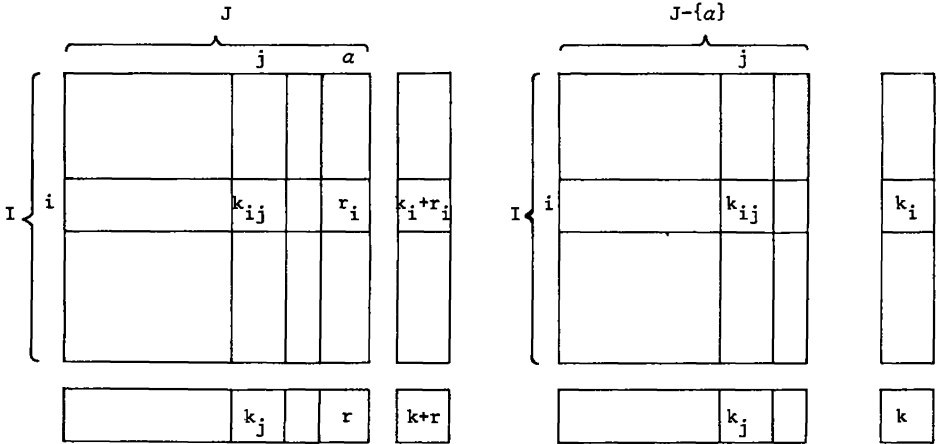


TABLEAU $I \times J$

TABLEAU $I \times (J - \{a\})$

où on a noté :

$$r_i = k_{ia} \quad k_j = \sum \{k_{ij} \mid i \in I\} \quad k_i = \sum \{k_{ij} \mid j \in J - \{a\}\}$$

$$r = \sum \{r_i \mid i \in I\} \quad k = \sum \{k_i \mid i \in I\}$$

$$f_i = (k_i + r_i) / (k + r)$$

Les facteurs sur I de la correspondance $I \times J$ sont notés $F_1, F_2, \dots, F_s, \dots$, leur inertie $\lambda_1, \lambda_2, \dots, \lambda_s, \dots$; ceux de la correspondance

$I \times (J - \{a\})$ sont notés $\hat{F}_1, \hat{F}_2, \dots, \hat{F}_S, \dots$ la valeur du facteur F_S en i est notée F_S^i ou $F_S(i)$.

2.1. Angle entre facteurs :

Les facteurs sur I qui sont dans les deux cas des fonctions sur I , sont donc des éléments de R^I . La correspondance $I \times J$ induit sur R^I le produit scalaire $\delta_i^{i'} f_i$. Il est naturel de prendre comme mesure de l'écart entre deux facteurs F et \hat{F} , leur angle θ dans cet espace R^I , défini par son cosinus :

$$\cos \theta = |\langle F, \hat{F} \rangle| / (\langle F, F \rangle \times \langle \hat{F}, \hat{F} \rangle)^{1/2}$$

où on a noté $\langle F, G \rangle = \sum \{F^i G^i f_i \mid i \in I\}$ le produit scalaire sur R^I associé à la mesure f_I .

Les facteurs F et \hat{F} ne sont pas orientés, leur angle est un angle de droite compris entre 0 et $\pi/2$, $\cos \theta$ est donc compris entre 0 et 1. Si F et \hat{F} sont identiques, θ est nul et $\cos \theta = 1$; au contraire si F et \hat{F} sont très différents, θ est proche de $\pi/2$ et $\cos \theta$ de 0. On sait en particulier que l'angle entre deux facteurs distincts de la correspondance $I \times J$ vaut $\pi/2$.

Remarque 1.

On peut interpréter $\cos \theta$ comme un coefficient de corrélation entre F et \hat{F} pour la mesure f_I , si les fréquences marginales des deux correspondances $I \times J$ et $I \times (J - \{a\})$ sont égales. En effet, dans ce cas les facteurs F et \hat{F} sont tous deux centrés par rapport à la mesure f_I :

$$\sum \{f_i F^i \mid i \in I\} = \sum \{f_i \hat{F}^i \mid i \in I\} = 0$$

Remarque 2.

La connaissance de θ permet aussi d'estimer approximativement la différence entre les positions respectives d'un élément i (autre que a) sur les facteurs F et \hat{F} , c'est-à-dire la valeur de $|F^i - \hat{F}^i|$ car on a approximativement :

$$\sum \{f_i (F^i - \hat{F}^i)^2 \mid i \in I\} \approx 2\lambda(1 - \cos \theta)$$

où λ est la valeur propre associée à F . (C'est une égalité si $\langle \hat{F}, \hat{F} \rangle = \langle F, F \rangle = \lambda$). Donc en moyenne, $|F^i - \hat{F}^i|^2 \approx 2\lambda(1 - \cos \theta)$. On montre aussi que la quantité $|F^i - \hat{F}^i|^2$ est très inférieure à cette valeur moyenne si i est proche du centre de gravité ou quand la contribution relative du facteur F à l'élément i est grande.

2.2. Angle entre plans factoriels :

Comparer les facteurs deux à deux est quelquefois insuffisant (cf. § 1 cas n° 2), on veut alors comparer des plans engendrés par deux facteurs, (ou même deux sous espaces de dimension supérieure à 2). On prendra alors comme mesure de l'écart entre ces plans l'angle maximum dans R^I (muni de la métrique associée à la mesure f_I) entre un vecteur de l'un de ces plans et sa projection sur l'autre plan. Cet angle est compris entre 0 et $\pi/2$; s'il vaut zéro, les plans sont confondus; s'il

vaut $\pi/2$, ils sont orthogonaux (i.e. le plan P contient une droite Δ orthogonale au plan P' et le plan P' contient une droite Δ' orthogonale à P). Cette définition se généralise à des sous espaces de dimension supérieure ([1], [Red. tens.], B n° 6 § 7).

Il ne nous appartient pas de donner, pour $\cos\theta$, un seuil à partir duquel on considérerait qu'un facteur ou un plan factoriel est stable; mais on peut remarquer que, dès que θ est inférieur à $\pi/4$, les facteurs comparés sont plus proches entre eux que de n'importe quel autre.

3. Comparaison des facteurs de l'analyse du tableau complet et du "tableau sans a".

Quand on retire de J, l'élément a, il se produit au niveau du nuage $\mathcal{N}(J)$ dans R_I deux modifications : d'une part, la métrique définie sur R_I est changée (elle était $(k+r)\delta_1^{i'}/(k_1+r_1)$, elle devient $k\delta_1^{i'}/k_1$), d'autre part, la forme quadratique d'inertie et le centre de gravité du nuage varient. Nous en étudions successivement les effets sur les facteurs.

3.1. Influence de la modification de la métrique.

La métrique étant souvent peu modifiée par la suppression d'un élément, cette influence est presque toujours négligeable. Si elle ne l'est pas, il est vraisemblable que les facteurs des deux analyses sont très différents, et il est inutile de poursuivre l'étude de l'influence de a. Cependant, si la rotation θ_M d'un facteur ou d'un plan factoriel provoquée par la modification de la métrique risque d'être faible mais non négligeable, il faut la prendre en compte dans la poursuite de l'étude de l'influence de a : nous avons montré qu'il faut ajouter la majoration de $\sin^2\theta_M$ à celle obtenue ensuite au niveau de la variation de l'inertie, pour obtenir une majoration de sinus de l'angle entre les facteurs (ou les plans factoriels) de même rang des correspondances $I \times J$ et $I \times J - \{a\}$.

Pour vérifier que cette influence sur le facteur de rang s est faible, c'est-à-dire que la rotation θ_M de F_s est petite, on utilise la majoration de $\sin 2\theta_M$ démontrée dans [4].

$$\sin 2\theta_M \leq \lambda_s \left(\sup_i (r_i/k_i) - \inf_i (r_i/k_i) \right) / \varepsilon (1 + \inf_i (r_i/k_i))$$

où on note :

$$\lambda_s = \text{valeur propre associée à } F_s$$

$$\varepsilon = \inf \{ \lambda_{s-1} - \lambda_s, \lambda_s - \lambda_{s+1} \}$$

Si deux valeurs propres consécutives λ_s et λ_{s+1} sont très proches, c'est le plan $[F_s, F_{s+1}]$ qu'il est nécessaire d'étudier; et nous avons montré que l'angle de rotation de ce plan est majoré de manière analogue en posant :

$$\varepsilon = \inf \{ (\lambda_{s-1} - \lambda_s), (\lambda_{s+1} - \lambda_{s+2}) \} / 2$$

3.2. Influence de la variation de l'inertie et du centre de gravité.

Supposons écartée l'influence due à la variation de la métrique (cf. § 3.1), et étudions maintenant l'influence de la variation de l'inertie et du centre de gravité.

Nous allons essayer de voir si la suppression de l'élément a a des conséquences proches de l'une ou de l'autre des situations présentées dans le § 1.

3.2.a. Le s-ième facteur F_s est-il stable ?

Sauf pour le premier facteur, nous supposons que le sous espace $[F_1, \dots, F_{s-1}]$ engendré par les facteurs de rang inférieur à celui du s-ème facteur est stable. Cette hypothèse est peu contraignante car généralement l'étude de la stabilité de F_s n'a d'intérêt que dans ce cas. Nous pouvons vérifier directement la stabilité de ce sous-espace (cf. remarque 1), mais nous conseillons d'étudier les facteurs (ou plan de deux facteurs successifs) dans l'ordre de leur inertie décroissante et de n'entreprendre l'étude de F_s que lorsque les facteurs (ou plans factoriels) de rang inférieur sont stables ce qui entraîne, ipso facto, la vérification de l'hypothèse de stabilité du sous espace $[F_1, \dots, F_{s-1}]$.

En toute rigueur, les majorations de l'angle θ entre les facteurs F_s et \hat{F}_s , que nous donnons ci-après, ne sont exactes que si le sous espace $[F_1, \dots, F_{s-1}]$ est absolument inchangé, mais elles ont une valeur approchée satisfaisante si on peut en négliger la rotation. (Cf. remarque 2).

Nous avons montré, sous cette hypothèse, une certaine stabilité du facteur F_s ($\theta < \pi/4$), si la condition (1) suivante est réalisée :

$$(1) \quad \lambda_s - ((k+r)/k) \text{CTR}_s(a) > \lambda_{s+1}$$

où $\text{CTR}_s(a) = (r/(k+r)) F_s^2(a)$ désigne la contribution absolue (*) de a au facteur F_s .

(Ceci signifie que l'inertie du nuage $\mathcal{N}(J - \{a\})$ projeté sur F_s reste supérieure à celle du nuage projeté sur F_{s+1} ; la suppression de a diminue l'inertie du nuage projeté sur F_s de $\text{CTR}_s(a)$, le coefficient $(k+r)/k$ provient du déplacement du centre de gravité).

Dans ce cas, nous savons déjà que θ est inférieur à $\pi/4$, et nous avons, de plus, plusieurs majorations de θ en fonction :

- du paramètre $m = (k+r) I_s(a)/k(\lambda_s - \lambda_{s+1})$ où $I_s(a)$ désigne, dans le nuage $\mathcal{N}(J)$, l'inertie de la projection de a sur le sous-espace engendré par les facteurs de rang supérieur ou égal à s : pour l'étude du premier facteur, $I_1(a)$ est l'inertie totale de a (qui se calcule facilement puisque les listages donnent en millièmes le rapport entre $I_1(a)$ et l'inertie totale du nuage ou somme des valeurs propres); pour le deuxième, $I_2(a)$ vaut $I_1(a)$ diminué de $\text{CTR}_1(a)$; pour le troisième, $I_3(a)$ vaut $I_2(a)$ diminué de $\text{CTR}_2(a)$, etc ...;

- de la contribution relative du facteur F_s à l'élément a notée $\cos^2 \phi_s (= f_a F_s^2(a)/I_1(a))$ qui est donnée dans les listages.

La majoration la plus simple de l'angle θ entre les facteurs F_s et \hat{F}_s n'est fonction que de m :

(*) Les listages donnent directement $(r/(k+r))F_s^2(a)/\lambda_s$ exprimé en millièmes.

$$\sin 2\theta \leq m \quad (\text{courbe III})$$

Les deux suivantes, plus fines, sont fonctions de la contribution relative de F_s à α , $\cos^2 \phi_s$:

$$\text{si } m \geq 1 \quad \text{tg } 2\theta \leq m \sin 2\phi_s / (1 - m \cos^2 \phi_s) \quad (\text{faisceau I})$$

$$\text{si } m < 1 \quad \text{tg } 2\theta \leq m \sin 2\phi_s / (1 - m \cos 2\phi_s) \quad (\text{faisceau II})$$

Pour faciliter l'utilisation de ces résultats, nous avons tracé des courbes représentant graphiquement $\cos \theta$: en fonction de m pour la première (courbe III); et pour les autres, en fonction de $\cos^2 \phi_s$ pour plusieurs valeurs du paramètre m (faisceau I et II).

Remarque 1 :

Sur l'hypothèse de stabilité du sous-espace $[F_1, \dots, F_{s-1}]$. On peut vérifier directement la stabilité de ce sous-espace en utilisant l'une ou l'autre des majorations de son angle de rotation θ :

$$\text{- en fonction de } m = (k+r) I_1(\alpha) / k(\lambda_{s-1} - \lambda_s) \quad (\text{cf. courbe III})$$

$$\text{- en fonction de } m \text{ et de } \cos^2 \psi = \cos^2 \phi_1 + \dots + \cos^2 \phi_{s-1}$$

contribution relative du sous-espace $[F_1, \dots, F_{s-1}]$ à α (cf. faisceau I si $m \geq 1$ et faisceau II si $m < 1$).

Sans vérifier cette hypothèse, on peut encore étudier la stabilité de F_s , car on a toujours :

$$\sin 2\theta \leq (k+r) \overline{I_1}(\alpha) / k \inf \{ \lambda_{s-1} - \lambda_s, \lambda_s - \lambda_{s+1} \}.$$

Remarque 2 :

Soit ψ l'angle de rotation du sous-espace F_1, \dots, F_{s-1} et θ_0 la majoration de θ obtenue en considérant que $\psi = 0$. On a :

$$\sin \theta \leq \sin \theta_0 + \sin \theta_R$$

$$\text{avec } \sin 2\theta_R \leq \frac{(\lambda_s - I_1(\alpha) + I_s(\alpha))\sqrt{2} \sin \psi + \lambda_1 \sin^2 \psi}{\lambda_s - \lambda_{s+1} - I_1(\alpha) + I_s(\alpha)}$$

En pratique, si ψ est petit, il est inutile de calculer cette majoration très large.

Remarque 3 :

Si la condition (1) n'est pas vérifiée, il peut y avoir encore, dans certains cas très particuliers une certaine stabilité de F_s , mais les méthodes exposées ici ne permettront pas de la reconnaître.

3.2.b. Le plan $[F_s, F_{s+1}]$ est-il stable ?

Si la stabilité de F_s n'a pu être démontrée, et si $\lambda_s - \lambda_{s+1}$ est très inférieur à $\lambda_{s+1} - \lambda_{s+2}$, il y a peut-être stabilité du plan $[F_s, F_{s+1}]$. Cette étude se fait de manière analogue à celle de F_s et nous supposons encore que le sous-espace $[F_1, \dots, F_{s-1}]$ peut être considéré comme inchangé.

Il y a une certaine stabilité du plan ($\theta < \pi/4$) si la condition (2) suivante est réalisée :

$$(2) \quad \lambda_{s+1} - ((k+r)/k) (CTR_s(a) + CTR_{s+1}(a)) > \lambda_{s+2}$$

(i.e. l'inertie du nuage projeté sur une droite quelconque de ce plan reste supérieure à celle du nuage projeté sur un facteur de rang supérieur à $s+1$). Les trois formules du paragraphe précédent sont valables si on prend :

$$- m = (k+r) I_s(a)/k(\lambda_{s+1} - \lambda_{s+2})$$

- et la contribution relative $\cos^2 \psi = \cos^2 \phi_s + \cos^2 \phi_{s+1}$ du plan $[F_s, F_{s+1}]$ à l'élément a .

Si le plan $[F_s, F_{s+1}]$ est très stable, on peut même préciser l'angle ρ de rotation de F_s dans ce plan : (ρ est l'angle entre F_s et \hat{F}_s). Car on a :

$$\tan 2\rho = m' \sin 2\psi' / (1 - m' \cos 2\psi')$$

où on a noté :

$$m' = (k+r) (CTR_s(a) + CTR_{s+1}(a)) / k(\lambda_s - \lambda_{s+1})$$

$$\cos^2 \psi' = \cos^2 \phi_s / (\cos^2 \phi_s + \cos^2 \phi_{s+1})$$

Le faisceau IV représente $\cos \rho$ en fonction de $\cos^2 \psi'$ pour quelques valeurs de m' . (i.e. on a avec les notations du § IV : $\rho = \theta_{II}(m', \psi')$).

3.2.c. Le facteur F_s disparaît-il ? se réinsère-t-il à un autre rang ?

Si, dans l'étude de l'influence de a , on n'a pu démontrer ni la stabilité de F_s , ni celle du sous-espace $[F_s, F_{s+1}]$ et si, de plus, la contribution relative $\cos^2 \phi_s$ de F_s à l'élément a est grande, il y a peut-être disparition de F_s ou décalage du rang de F_s (cf. § I-3 et I-4).

Nos méthodes permettent d'abord de connaître le rang t ($t > s$) auquel le facteur F_s peut se réinsérer.

- Si ce rang est trop grand pour que le facteur reste parmi les facteurs interprétables, il peut y avoir disparition du facteur F_s . Nos méthodes permettent alors de le montrer : i.e. de montrer que les facteurs de rang strictement inférieur à s de la correspondance $I \times J$ sont encore facteurs de $I \times (J - \{a\})$ avec le même rang; et que les facteurs interprétables de rang strictement supérieur à s sont encore facteurs de $I \times J - \{a\}$, leur rang ayant diminué de 1. (F_{s+1}, F_{s+2}, \dots , respectivement proches de $\hat{F}_s, \hat{F}_{s+1}, \dots$).

- Si ce rang t est tel que le facteur peut rester parmi les facteurs interprétables, il peut y avoir réinsertion du facteur F_s au rang t . Quand elle a lieu, nos méthodes nous permettent de montrer que le facteur F_s reste facteur de $I \times (J - \{a\})$ mais avec le rang t . Elles nous permettent aussi d'étudier les facteurs suivants : de comparer F_{s+1} à F_s, F_{s+2} à F_{s+1}, \dots, F_t à F_{t-1} ; et de comparer les autres facteurs interprétables de même rang (supérieur à t) des deux analyses. Ces comparaisons sont indépendantes entre elles. Dans certains cas, on pourra montrer la conservation de certains facteurs (dont peut-être F_s) et on ne pourra rien conclure pour d'autres.

Pour connaître le rang t , auquel F_s peut être décalé, on calcule la quantité :

$$\lambda_s - (k+r)/k I_1(a) \cos 2\phi_s$$

- Si cette quantité reste supérieure à λ_{s+1} , on ne pourra rien conclure; vraisemblablement les facteurs de rang supérieur ou égal à s de $I \times J - \{a\}$ sont très différents dans les deux analyses.
- Si cette quantité est comprise entre 2 valeurs propres λ_t et λ_{t+1} avec $t \geq s+1$, il peut y avoir décalage de F_s au rang t , ou disparition de F_s si t est trop grand. On a représenté cette situation dans les tableaux ci-dessous. Les facteurs de $I \times J$ et de $I \times (J - \{a\})$ à comparer

TABLEAU I.

$t > s + 1$: décalage possible du facteur F_s au rang t .

Facteurs de $I \times (J - \{a\})$	Facteurs de $I \times J$ à comparer	Valeurs propres associées
⋮	⋮	⋮
\hat{F}_s	F_{s+1}	$\lambda'_s = \lambda_{s+1}$
\hat{F}_{s+1}	F_{s+2}	$\lambda'_{s+1} = \lambda_{s+2}$
⋮	⋮	⋮
\hat{F}_t	F_s	$\lambda'_t = \lambda_s - ((k+r)/k) I_1(a) \cos 2\phi_s$
\hat{F}_{t+1}	F_{t+1}	$\lambda'_{t+1} = \lambda_{t+1}$
⋮	⋮	⋮
⋮	⋮	⋮

TABLEAU II.

$t = s + 1$: permutation possible de F_s et F_{s+1} (voir exemple § 5.2).

Facteurs de $I \times (J - \{a\})$	Facteurs de $I \times J$ à comparer	Valeurs propres associées
⋮	⋮	⋮
\hat{F}_s	F_{s+1}	$\lambda'_s = \lambda_{s+1}$
\hat{F}_{s+1}	F_s	$\lambda'_{s+1} = \lambda_s - ((k+r)/k) I_1(a) \cos 2\phi_s$
\hat{F}_{s+2}	F_{s+2}	λ_{s+2}
⋮	⋮	⋮
⋮	⋮	⋮
⋮	⋮	⋮

sont écrits sur une même ligne et afin de permettre l'écriture d'une formule unique de la majoration de l'angle entre deux facteurs, nous avons noté λ'_u la valeur propre associée au facteur de $I \times J$ que l'on compare à F_u ; pour F_s cette valeur est modifiée en $\lambda'_t = \lambda'_s - ((k+r)/k) \times I_1(a) \cos 2\phi_s$ (la suite des λ'_u est décroissante). Le tableau I représente le cas où $t > s + 1$. Si $t = s + 1$, il peut y avoir permutation de F_s et de F_{s+1} , ce cas est représenté au tableau II.

Nous avons montré que l'angle θ entre \hat{F}_u et le facteur de $I \times J$ qu'on lui compare est majoré par :

$$\sin 2\theta \leq m = (k+r) I_1(a) \sin 2\phi_s / k \inf \{ \lambda'_u - 1 - \lambda'_u, \lambda'_u - \lambda_{u+1} \}.$$

La courbe III donne, en fonction de m la minoration de $\cos \theta$.

Nous pouvons comparer aussi, si besoin est, deux plans factoriels : l'angle θ entre le plan $[\hat{F}_u, \hat{F}_{u+1}]$ et le plan analogue de $I \times J$ est majoré de la même façon en remplaçant le dénominateur de m par :

$$k \inf \{ \lambda'_u - 1 - \lambda'_u, \lambda'_{u+1} - \lambda'_{u+2} \}.$$

Remarque 1.

Il n'est pas nécessaire ici de supposer la stabilité du sous espace $[F_1, \dots, F_{s-1}]$ engendré par les facteurs de rang inférieur à s . Cependant, si ce sous-espace est stable, on peut diminuer la borne en remplaçant $I_1(a)$ par $I_s(a)$ et ϕ_s par ψ_s avec $\cos^2 \psi_s = \cos^2 \phi_s / (1 - (\cos^2 \phi_1 + \dots + \cos^2 \phi_{s-1}))$. La borne est diminuée si la contribution relative de a au sous-espace $[F_1, \dots, F_{s-1}]$ est assez importante.

Remarque 2.

- Cette majoration, valable quel que soit u , s'applique à l'étude des facteurs de rang inférieur à s . Ceux-ci peuvent donc être étudiés aussi bien par les méthodes du § 3.2.a et b que par celle de ce §.

- Par contre, dans l'étude des facteurs de rang s jusqu'à t , seules les majorations de ce § s'appliquent, puisque les facteurs à comparer ne sont pas de même rang.

- De même, on peut utiliser pour les facteurs de rang supérieur à t la majoration de ce §, ou, si le sous espace $[F_1, \dots, F_t]$ est stable celle des § 3.2.a et § 3.2.b.

Remarque 3.

Dans l'étude du décalage de F_s au rang t , c'est en utilisant les valeurs propres λ'_u qu'il convient d'étudier l'influence due au changement de métrique. (Cf. § 3.1).

4. Résumé-guide de l'utilisateur :

Les résultats cités dans le paragraphe précédent, montrent que les éléments menaçant la stabilité des résultats d'une analyse sont ceux qui ont une contribution absolue importante pour un ou plusieurs facteurs; et que de plus les facteurs les plus menacés sont ceux qui ont une valeur propre peu séparée des autres. Ces éléments sont très facilement repérables dans les résultats fournis par les programmes d'analyse des correspondances (qui donnent, pour chaque facteur, le quotient de la contribution absolue par la valeur propre (i.e. $f_i F_s^2(i)/\lambda_s$), exprimé en millièmes).

Pour l'étude d'un tel élément, il est conseillé de considérer les facteurs dans l'ordre de leur inertie décroissante, et de n'entreprendre l'étude d'un facteur que si les précédents (ou les plans factoriels précédents) sont assez stables. (cf. § 3.2.a début et remarque 1). Il est conseillé aussi, de suivre, pour chaque facteur, le plan d'étude proposé : tenter d'appliquer le § 4.1, puis, si cela n'est pas possible, tenter d'appliquer 4.2, si 4.2 ne s'applique pas non plus, tenter d'appliquer 4.3. Si aucun de ces 3 paragraphes ne s'applique, on ne pourra rien conclure.

Rappel des notations (cf. § 2)

$$k = \Sigma \{k_{ij} | i \in I, j \in J - \{a\}\}$$

$$r_i = k_{ia} \quad r = \Sigma \{r_i | i \in I\}$$

$$k+r = \Sigma \{k_{ij} | i \in I, j \in J\}$$

F_s facteur de rang s de la correspondance $I \times J$

\hat{F}_s facteur de rang s de la correspondance $I \times (J - \{a\})$

λ_s valeur propre associée à F_s

$\cos^2 \phi_s$ contribution relative de F_s à a

$CTR_s(a) = (r/(k+r)) F_s^2(a)$ contribution absolue de a à F_s

$I_s(a)$ inertie de a sur le sous-espace engendré par les facteurs de rang supérieur ou égal à s : $I_1(a) = CTR_1(a)/\cos^2 \phi_1$

$$I_{s+1}(a) = I_s(a) - CTR_s(a)$$

θ angle entre F_s et \hat{F}_s ou (étude d'un plan) angle entre $[F_s, F_{s+1}]$ et $[\hat{F}_s, \hat{F}_{s+1}]$.

Influence de la métrique (cf. § 3.1).

$$\sin 2\theta_M \leq \lambda_s (\sup_i (r_i/k_i) - \inf_i (r_i/k_i)) / \epsilon (1 + (r_i/k_i))$$

où $\epsilon = \inf \{ \lambda_s - 1 - \lambda_s, \lambda_s - \lambda_{s+1} \}$ pour l'étude de F_s

et $\epsilon = \inf \{ \lambda_s - 1 - \lambda_s, \lambda_{s+1} - \lambda_{s+2} \} / 2$ pour l'étude du plan $[F_s, F_{s+1}]$

4.1. si $\lambda_s - \lambda_{s+1} - (k+r)CTR_s(a)/k > 0$: stabilité de F_s

Après avoir vérifié que l'influence de la métrique est négligeable, on sait que $\theta < \pi/4$. Il faut en chercher une borne plus précise. Calculer :

$$m = (k+r) I_s(a) / k(\lambda_s - \lambda_{s+1})$$

- si m est très petit, F_s est stable. La courbe III suffit.
- si m est assez grand, mais inférieur à 1, F_s est encore stable si $\cos^2 \phi_s$ est proche de 0 ou très proche de 1. Utiliser le faisceau II.
- si m est supérieur à 1, F_s est encore stable si $\cos^2 \phi_s$ est proche de 0 (cf. faisceau I).

4.2. Si $\lambda_{\delta+1} - \lambda_{\delta+2} - (k+r)(CTR_{\delta}(a) + CTR_{\delta+1}(a))/k > 0$ - Stabilité du plan $[F_{\delta}, F_{\delta+1}]$.

Etude analogue à celle de 4.1. Vérifier que l'influence de la métrique est négligeable. Puis calculer :

$$m = (k+r) I_s(a) / k(\lambda_{s+1} - \lambda_s)$$

- Si m est petit le plan est stable (courbe III)
- Si $m < 1$, le plan est stable si $\cos^2 \psi = \cos^2 \phi_s + \cos^2 \phi_{s+1}$ est proche de 0 ou très proche de 1 (faisceau II)
- Si $m > 1$, faisceau I.

Si le plan est stable, calculer la rotation ρ de F_s dans ce plan avec le faisceau IV en fonction de $\cos^2 \psi' = \cos^2 \phi_s / (\cos^2 \phi_s + \cos^2 \phi_{s+1})$.

4.3. Si $\lambda_t > \lambda_{\delta} - ((k+r)/k) I_1(a) \cos^2 \phi_{\delta} > \lambda_{t+1}$ avec $t \geq s+1$.

Décalage de F_s au rang t ou disparition de F_s .

Dans ce cas, on compare les facteurs de même rang des deux analyses sauf ceux dont le rang est compris entre s et t : i.e. on compare \hat{F}_1 à $F_1, \dots, \hat{F}_{s-1}$ à $F_{s-1}; \hat{F}_s$ à $F_{s+1}, \dots, \hat{F}_t - 1$ à $F_t; \hat{F}_t$ à $F_s; \hat{F}_t + 1$ à F_{t+1} etc...; on note λ'_u la valeur propre associée au facteur de $I \times J$ comparé à F_u corrigée pour F_s en :

$$\lambda'_t = \lambda'_s - (k+r) I_1(a) \cos^2 \phi_s / k \quad (\text{cf. tableau p.304})$$

- La minoration du cosinus de l'angle θ entre \hat{F}_u et le facteur de $I \times J$ auquel on le compare se lit sur la courbe III en fonction du paramètre $m = (k+r) I_1(a) \sin^2 \phi_s / k \inf (\lambda'_u - 1 - \lambda'_u, \lambda'_u - \lambda'_{u+1})$

- Si le sous-espace $[F_1, \dots, F_{s-1}]$ est stable, on diminue m en remplaçant $I_1(a)$ par $I_s(a)$, ϕ_s par ψ_s tel que :

$$\cos^2 \psi_s = \cos^2 \phi_s / (1 - \cos^2 \phi_1, \dots, - \cos^2 \phi_{s-1}).$$

- Pour vérifier que l'influence de la métrique est négligeable, il faut utiliser les λ'_u et non les λ_u .

5. Exemple d'application.

Pour illustrer la méthode, nous avons construit, à partir de données réelles un tableau d'effectifs sur $I \times J$ avec card $I = 51$ et card $J = 41$. Nous donnons ici les résultats de l'analyse tels qu'ils apparaissent sur les listages : la table des valeurs propres et les six premiers facteurs.

Avant d'étudier l'influence de certains éléments, considérons la suite des valeurs propres. En effet, les résultats cités dans le para-

graphe précédent montrent qu'un facteur est d'autant plus stable que l'écart entre sa valeur propre et les autres valeurs propres est grand.

Les trois premières valeurs propres sont chacune bien séparée des autres; à partir de la quatrième, les valeurs propres décroissent beaucoup plus lentement. Les facteurs de rang supérieur à 3 risquent donc d'être plus sensibles aux modifications de données. De plus, $\lambda_4 - \lambda_5$ étant beaucoup plus petit que $\lambda_5 - \lambda_6$, on sera peut-être amené à considérer le plan $\{F_4, F_5\}$ a priori plus stable que le facteur F_4 .

Examinons maintenant les facteurs sur J. Un élément A attire particulièrement l'attention par l'importance de sa contribution à l'inertie du premier facteur (213‰). De plus sa contribution relative $\cos^2 \phi_1$ vaut 0,992. Le premier facteur sera-t-il modifié par la suppression de A ?

5.1. Etude de la suppression de A.

5.1.a. Le premier facteur.

Calculons $CTR_1(A) = \lambda_1 \times 0.213 = 0.0025$

Le quotient $(k+r)/k$ vaut ici 1.07.

On a $\lambda_1 - \lambda_2 - (k+r) CTR_1(A)/k > 0$. Nous pouvons donc considérer l'hypothèse de la stabilité du premier facteur.

Vérifions que la rotation du premier facteur due à la modification de la métrique est négligeable. Dans ce cas, on a :

$$\sup_i r_i/k_i = 0.112 \quad \text{et} \quad \inf_i r_i/k_i = 0.030$$

d'où $\sin 2\theta_M \leq 0.22$ et $\cos \theta_M > 0.994$. Cet effet est donc négligeable.

Etudions maintenant l'effet de la modification de l'inertie et du centre de gravité. On est déjà assuré que la rotation du premier facteur est inférieure à 45° . Calculons alors le paramètre $m = ((k+r)/k)I_1(A)/(\lambda_1 - \lambda_2)$: l'inertie de l'élément A par rapport au centre de gravité du nuage vaut 69% de l'inertie du nuage, c'est-à-dire : $0.069 \times (0.0379) = 0.0026$. On en déduit : $m = 0.65$. Cette valeur est inférieure à 1, la courbe III est utilisable et nous permet d'affirmer que $\cos \theta > 0.94$. Comme $\cos^2 \phi_1 (= 0.992)$ est presque égal à 1, le faisceau II permet d'affiner ce résultat : $\cos \theta > 0.99$.

La suppression de A ne modifie donc pas le premier facteur.

5.1.b. Etudions maintenant l'influence de A sur les facteurs suivants.

L'influence de la modification de l'inertie est quasiment nulle. On peut le montrer de deux manières différentes : soit en remarquant que tous les $\cos^2 \phi_s$ sont presque égaux à 0, ce qui assure la stabilité quel que soit le paramètre m (cf. § 3.2.a. ou faisceau II et III); soit en montrant que, pour ces facteurs, la valeur du paramètre m est très petite. En effet, $I_2(A)$ est l'inertie de A diminuée de son inertie sur le premier facteur : $I_2(A) = 0.0026 \times (1-0.992) = 0.00002$. Les écarts entre les valeurs propres sont tous supérieurs à 10 fois cette valeur, on a donc pour ces facteurs $m < 0.1$.

Voyons si l'influence de la modification de métrique est encore négligeable sur ces facteurs :

Facteur	2	3	4	5	Plan [4,5]
$\sin 2\theta_M <$	0.202	0.167	0.806	0.727	0.43
$\cos \theta_M >$	0.995	0.996	0.89	0.92	0.97

Pour le deuxième et le troisième facteur, la majoration de θ_M est encore très petite. Les 3 premiers facteurs sont donc stables.

Pour le quatrième et le cinquième facteur, la variation est encore petite, mais on ne peut plus affirmer qu'elle est négligeable. Cependant, comme les majorations de θ_M sont très larges, il est vraisemblable que les facteurs 4 et 5 varient également encore très peu. Si l'on considère globalement le plan qu'ils engendrent, on peut être encore plus affirmatif.

5.2. Etude de la suppression de B.

Ce sont la contribution absolue de B au deuxième facteur, et la contribution relative du deuxième facteur à B qui sont particulièrement importantes, c'est donc surtout ce deuxième facteur qui risque d'être modifié par la suppression de B.

Le premier facteur n'est pas modifié puisque $\cos^2 \phi_1$ est nul et que la modification de la métrique est négligeable. (On a :

$$\sup r_i/k_i = 0.05 \text{ et } \inf (r_i/k_i) = 0.009).$$

Etudions le deuxième facteur :

$$\text{CTR}_2(B) = \lambda_2 \times 0.498 = 0.0038$$

$$(k+r)/k = 1.02$$

il vient :

$$\lambda_2 - \lambda_3 - (k+r) \text{ CTR}_2(B)/k < 0.$$

Le deuxième facteur est vraisemblablement modifié. Il est inutile de tenter de démontrer la stabilité du plan $\{F_2, F_3\}$, puisque l'écart $\lambda_3 - \lambda_4$ est inférieur à $\lambda_2 - \lambda_3$. Par contre, on peut étudier un éventuel décalage ou une disparition de F_2 . Calculons donc :

$$\lambda_3' = \lambda_2 - ((k+r)/k) I_1(B) \cos^2 \phi_2 = 0.00379$$

$$\text{on a : } \lambda_1 > \lambda_3 > \lambda_3' > \lambda_4 > \dots$$

Puisque λ_3' est située entre λ_3 et λ_4 , il y a peut-être permutation de F_2 et de F_3 ; Nous indiquons dans le tableau ci-dessous les valeurs propres corrigées associées, ainsi que l'écart minimum ε entre cette valeur propre et les autres.

Facteur de I x J - {a}	Facteur de I x J	Valeur propre associée	ε
\hat{F}_1	F ₁	$\lambda_1^1 = \lambda_1 = 0.01194$	$\varepsilon_1 = \lambda_1 - \lambda_2 = 0.00432$
\hat{F}_2	F ₃	$\lambda_2^1 = \lambda_3 = 0.00462$	$\varepsilon_2 = \lambda_3 - \lambda_3^1 = 0.00084$
\hat{F}_3	F ₂	$\lambda_3^1 = 0.00379$	$\varepsilon_3 = \lambda_3 - \lambda_3^1 = 0.00084$
\hat{F}_4	F ₄	$\lambda_4^1 = \lambda_4 = 0.00263$	$\varepsilon_4 = \lambda_4 - \lambda_5 = 0.00024$

Comparons \hat{F}_2 et F₃.

Vérifions que la variation due à la modification de la métrique est négligeable :

$$\sin 2\theta_M \leq \lambda_3 \text{ (sur } (r_i/k_i) - \inf (r_i/k_i)/\varepsilon_2 (1 + \inf (r_i/k_i)) \leq 0.25$$

Calculons le paramètre m :

$$m = (k+r)I_1(B) \sin 2\phi_2/k \varepsilon_2 = 1.02 \times 0.00041/0.00083 = 0.5$$

La courbe III nous indique que $\cos \theta$ est supérieur à 0.96 : le deuxième facteur \hat{F}_2 de "l'analyse sans B" est donc proche du troisième facteur F₃ de l'analyse globale.

Comparons \hat{F}_3 et F₂.

L'influence de la métrique est encore négligeable puisque $\varepsilon_3 = \varepsilon_2$ et $\lambda_3^1 < \lambda_3$. La majoration de l'angle entre \hat{F}_3 et F₂ est identique à la précédente, la conclusion est identique. La suppression de B permute les facteurs F₂ et F₃.

Comparons \hat{F}_4 et F₄.

La valeur de m est supérieure à 1, on ne peut donc préciser par cette méthode la position du facteur \hat{F}_4 ni des suivants.

Par contre, on peut utiliser les méthodes du § 3.2.a. puisque le sous-espace [F₁, F₂, F₃] varie assez peu. On peut ainsi montrer la stabilité de F₄ et de F₅ puisque $\cos^2 \phi_4 = \cos^2 \phi_5 = 0$, et que l'influence de la métrique est négligeable.

5.3. Etude de la suppression de C

La contribution de C au quatrième facteur 120 °/° est assez importante et sa suppression risque de le modifier car l'écart entre λ_4 et λ_5 est très faible.

Pour la modification de l'inertie, on obtient les résultats suivants: $(k+r)/k = 1.025$.

Facteur	$CTR_s(C) \times (k+r) / k < \lambda_s - \lambda_{s+1}$	$I_s(C)$	m	$\cos \theta >$
F ₁	0.0012 oui	0.0019	0.45	0.97
F ₂	0.0001 oui	0.0006	0.2	0.99
F ₃	0.00007 oui	0.00056	0.25	0.99
F ₄	0.00029 non	0.00046		

D'autre part, les modifications dues à la métrique sont négligeables. Les 3 premiers facteurs sont donc stables. Le facteur F₄ ne l'est probablement pas, étudions le plan [F₄, F₅]. On a :

$\{CTR_4(C) - CTR_5(C)\} (k+r) / k < \lambda_5 - \lambda_6$	$I_4(C)$	m	$\cos^2 \psi = \cos^2 \phi_4 + \cos^2 \phi_5$	$\cos \theta >$
0.00030	0.00040	1.15	0.155	0.91

Ce plan a donc une certaine stabilité ($\sin \theta < 0,3$).

Pour préciser la rotation de F₄ dans ce plan, on utilise le faisceau IV qui pour $\cos^2 \psi = \cos^2 \phi_4 / (\cos^2 \phi_4 + \cos^2 \phi_5) = 0.987$ nous indique :

$$\cos \rho \approx 0.442 \quad (\rho = \theta_{II} (m, \psi) \approx 63^\circ)$$

E X E M P L E

Tableau des valeurs propres
et des facteurs

VAL PROPRE	POURCENT	CUMUL
.01193749	31.533	31.533
.00761607	20.118	51.650
.00461963	12.203	63.853
.00242895	6.416	70.269
.00218918	5.783	76.052
.00178639	4.719	80.770
.00107132	2.830	83.600
.00087963	2.324	85.924
.00074206	1.960	87.884
.00071143	1.879	89.763
.00056185	1.484	91.247
.00052204	1.379	92.626
.00046670	1.233	93.859
.00032886	.869	94.727
.00028935	.764	95.492
.00028023	.740	96.232

$$\lambda_1 - \lambda_2 = 0.00432$$

$$\lambda_2 - \lambda_3 = 0.00299$$

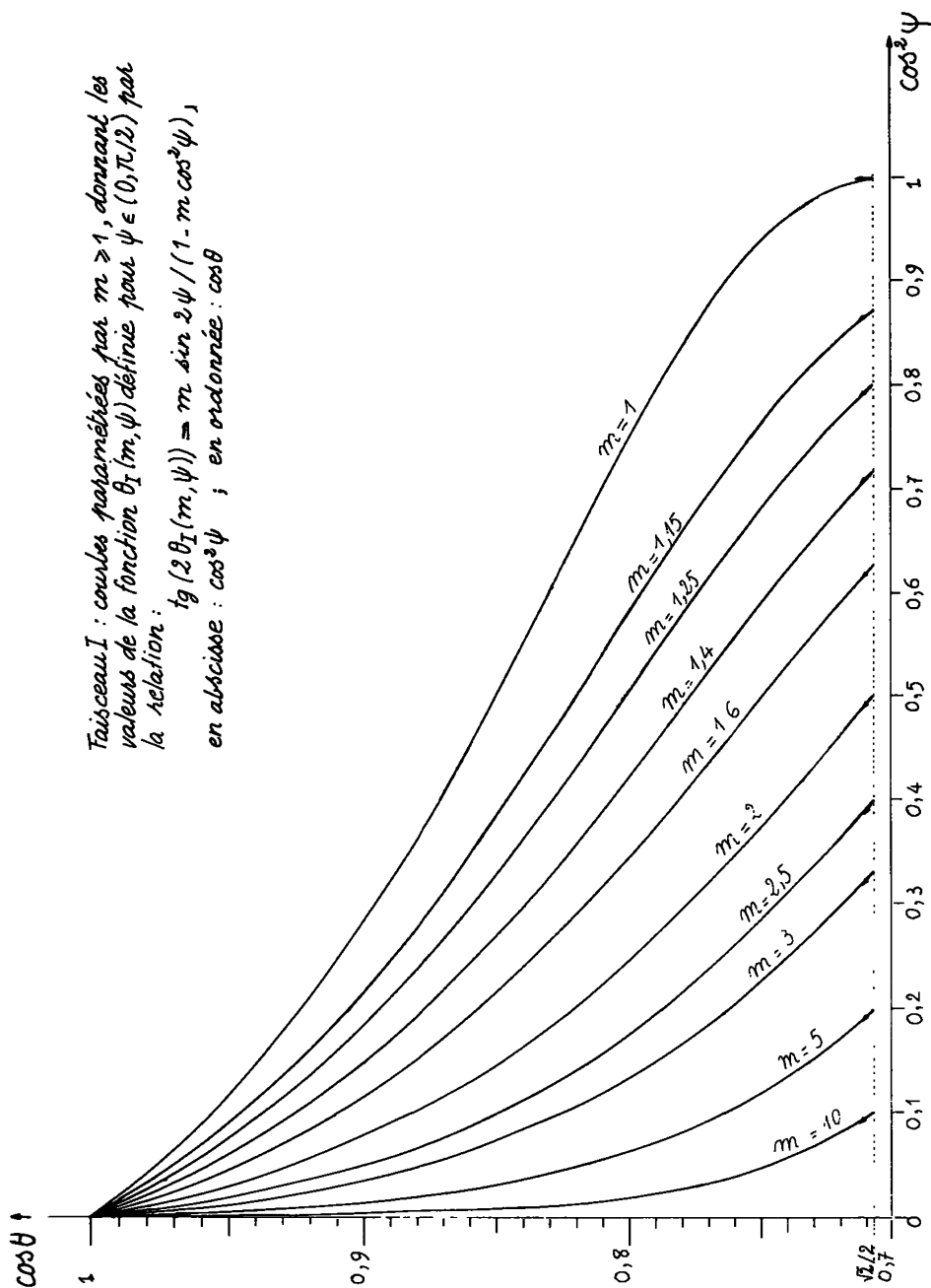
$$\lambda_3 - \lambda_4 = 0.00219$$

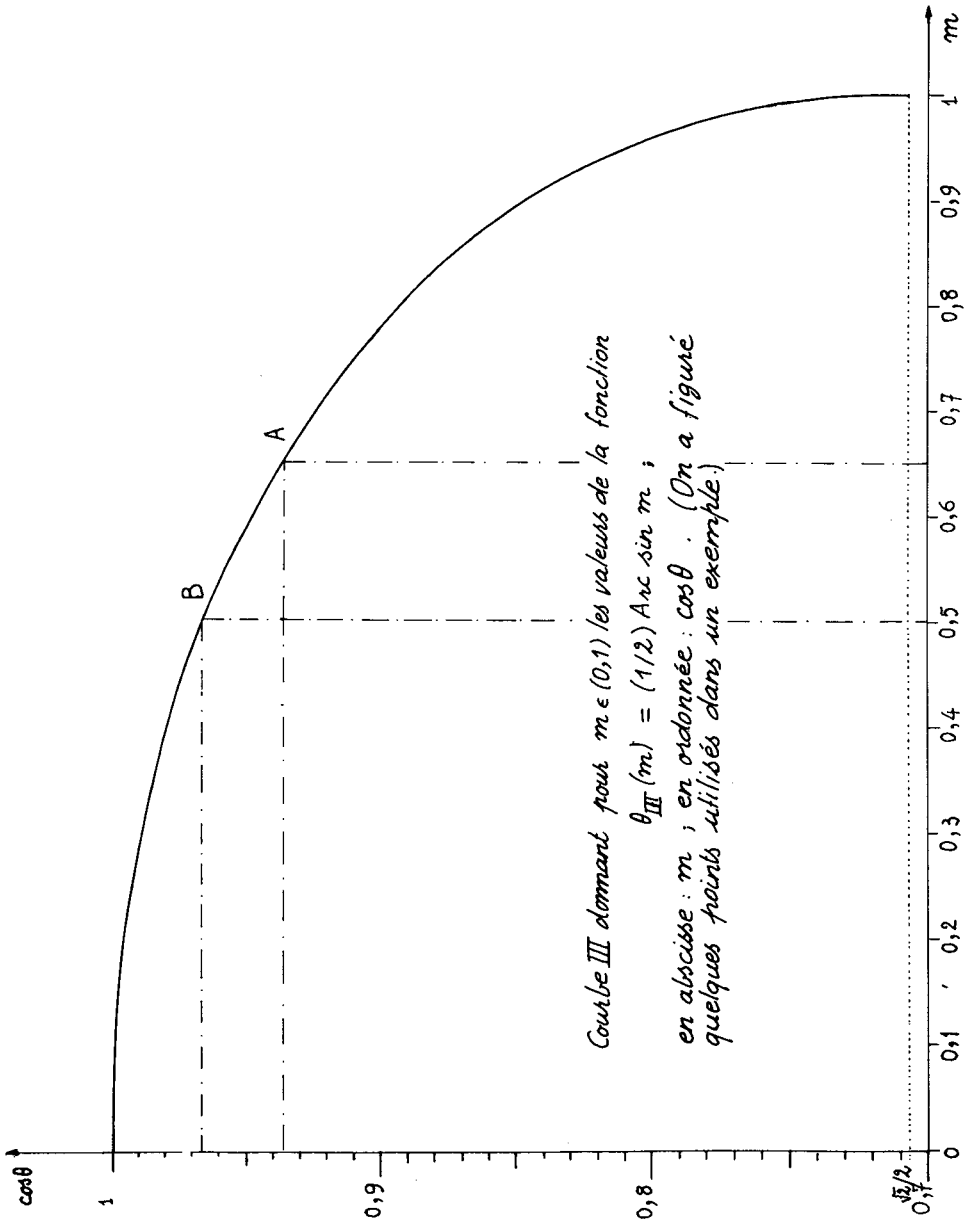
$$\lambda_4 - \lambda_5 = 0.00024$$

$$\lambda_5 - \lambda_6 = 0.00040$$

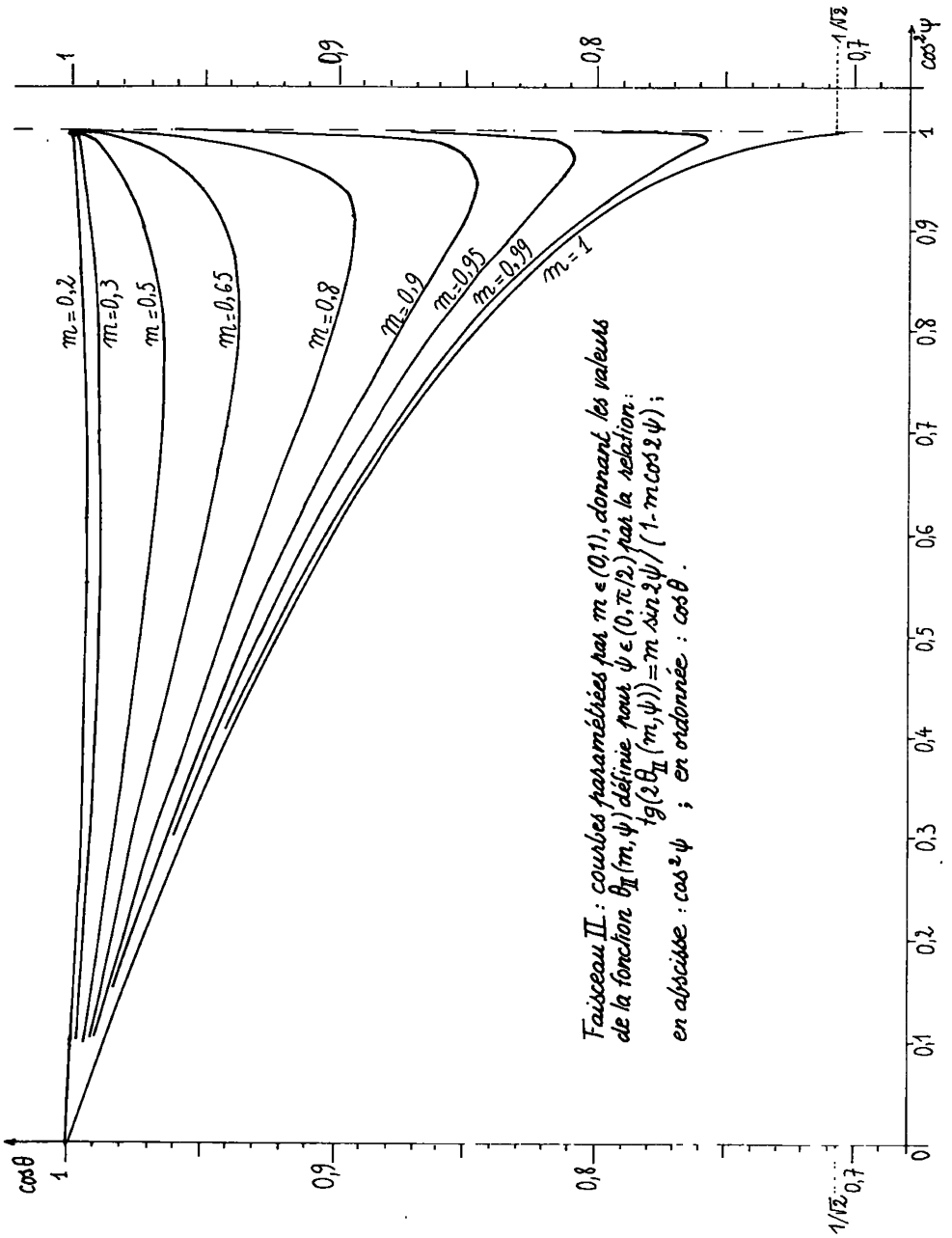
$$\lambda_6 - \lambda_7 = 0.00071$$

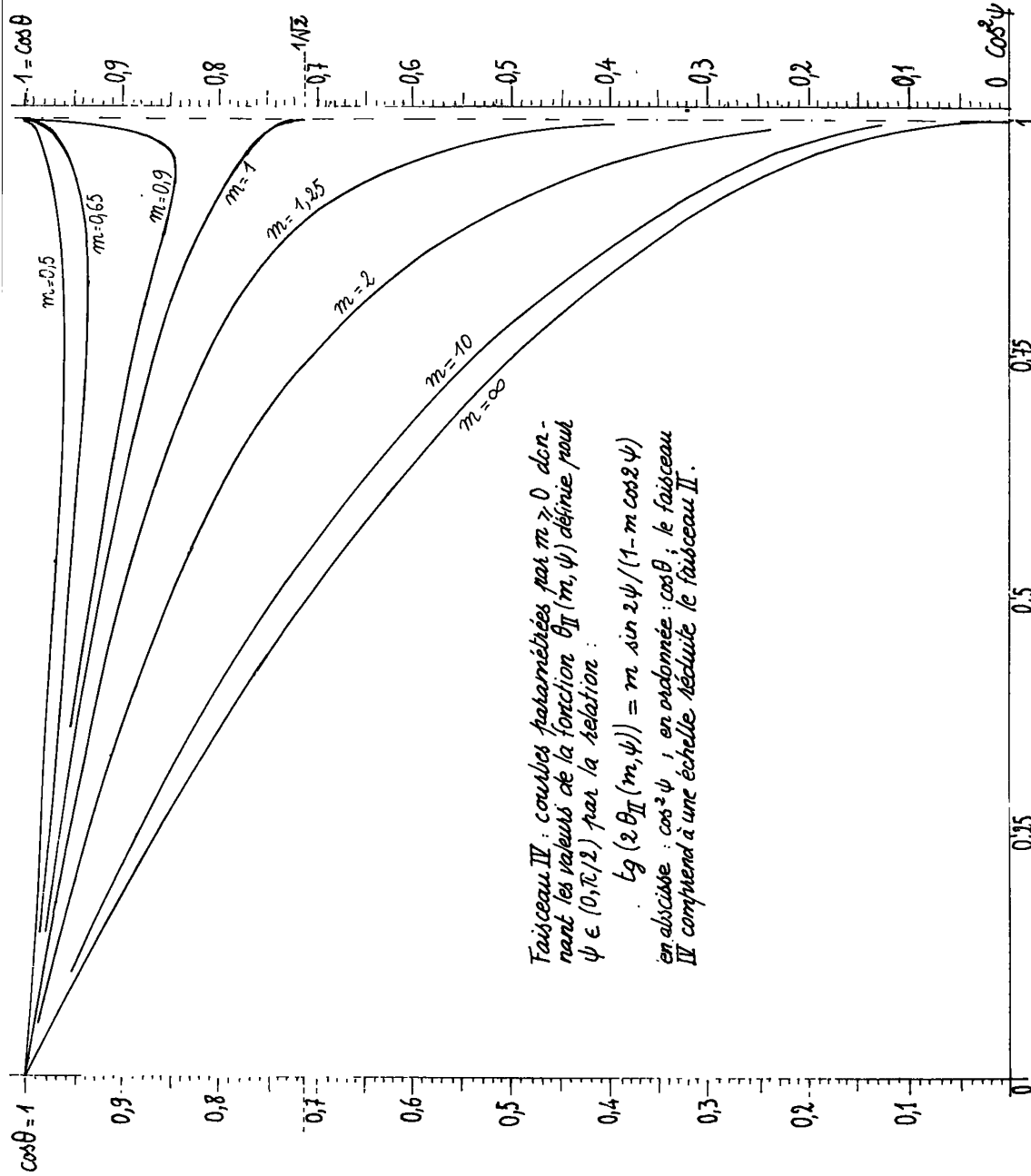
Taureau I : courbes paramétrées par $m \geq 1$, donnant les valeurs de la fonction $\beta_I(m, \psi)$ définie pour $\psi \in (0, \pi/2)$ par la relation : $\operatorname{tg}(2\beta_I(m, \psi)) = m \sin 2\psi / (1 - m \cos^2 \psi)$, en abscisse : $\cos^2 \psi$; en ordonnée : $\cos \theta$





Courbe III donnant pour $m \in (0,1)$ les valeurs de la fonction $\theta_{III}(m) = (1/2) \text{Arc sin } m$; en ordonnée : $\cos \theta$. (On a figuré quelques points utilisés dans un exemple.)





Faisceau IV : courbes paramétrées par $m \geq 0$ donnant les valeurs de la fonction $\theta_{II}(m, \psi)$ définie pour $\psi \in (0, \pi/2)$ par la relation :

$$\operatorname{tg}(2\theta_{II}(m, \psi)) = m \sin 2\psi / (1 - m \cos 2\psi)$$

en abscisse : $\cos^2 \psi$; en ordonnée : $\cos \theta$; le faisceau IV comprend à une échelle réduite le faisceau II.

B I B L I O G R A P H I E

- [1] BENZECRI J.P. et coll. L'analyse des données - T.II L'analyse des correspondances - Dunod - Paris 1973.
- [2] ESCOFIER B. et LE ROUX B. Etude des questionnaires par l'analyse des correspondances - Modification du codage des questions ou de leur nombre et stabilité de l'analyse. Math. et Sc. hum., n° 49 (1975).
- [3] ESCOFIER B. et LE ROUX B. Mesure de l'influence d'un descripteur sur les résultats d'une analyse en composantes principales - 1974. Pub. de l'inst. de Stat. (à paraître).
- [4] ESCOFIER B. et LE ROUX B. Ajout et suppression d'éléments en analyse des correspondances. Texte multigraphié. INSA Rennes.